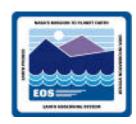


Reprocessing Scott H. Thompson

sthompso@eos.hitc.com

21 February 1996

Agenda



Head of Chain

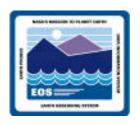
- Environment
- Input replication methods

Results

- PGS CPU
- Data Handler R/W Stations
- Data Handler Robots

Conclusions

What is examined?



August v2.2 baseline, including users and V0

All instruments, except MODIS L3/L4

Resource allocation

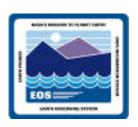
- Processing CPUs (<50% from Base run)
- Data handler r/w stations and robots (<50% from Base run)

Methodologies examined

- Base (no reprocessing)
- Base (with reprocessing through duplication + priority)
- Base (with reprocessing through frequency)

Analyze PGS CPUs, DH R/W Stations & Robots

What is Duplication + Priority?



This method produces a second copy of all process and file descriptions in a base run, thus doubling the processing load in the system to reflect reprocessing.

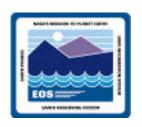
These newly created processes are given a priority indicative of their importance.

• In this case these new processes are given a lower priority than all normal (first-time) and user processing.

Normal and duplicated processes enter the system at the same time and compete for resources. The only determining factor on who obtains a resource is the assigned priority.

Note: Normal processing has a higher priority over user processing.

What is Frequency?

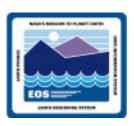


This method modifies the process and file descriptions so that each process executes twice as often per day, thus doubling the processing load in the system to reflect reprocessing.

As the name implies, Frequency produces a second set of processes at a 180° shift from the first set of processes. The second set can be considered to be the reprocessing jobs.

Competition

What was the PGS CPU usage?

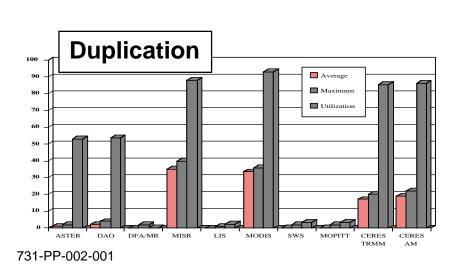


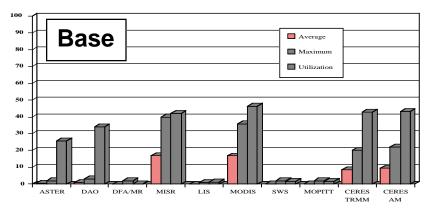
• Statistics

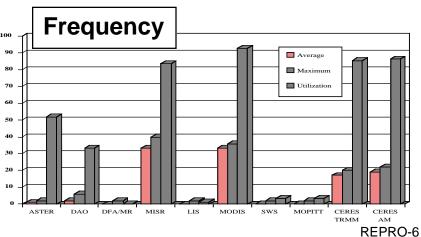
- Average use over 21 day run
- Maximum instance
 - » not always allocated amount

Utilization

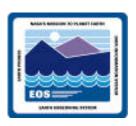
Average / Maximum







How did the PGS CPUs stack up?



Examining those instruments with more than 2 CPUs

• CERES TRMM, CERES AM, MISR, MODIS

CPU Utilization

• Base 42% to 47%

• Duplication 83% to 93%

• Frequency 85% to 93%

As expected reprocessing "doubled" the utilization for the PGS CPUs for each instrument

What was the DH R/W Station usage?

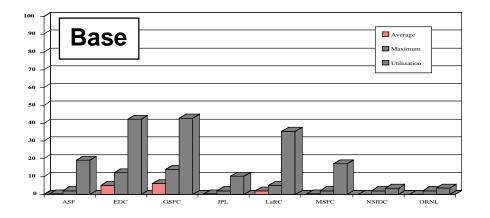


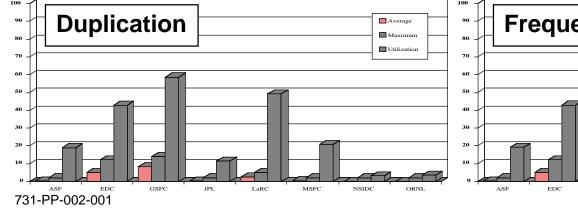
Statistics

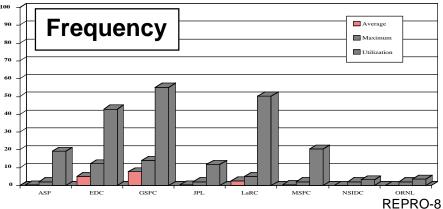
- Average use over 21 day run
- Maximum instance

Utilization

- Average / Maximum







How did the DH R/W Stations stack up?



Examined those sites with more than an average of ten DH R/W stations in use (Base run)

• EDC, GSFC, LaRC

Utilization

• Base 48% to 49%

• Duplication 50% to 72%

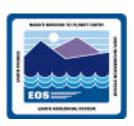
• Frequency 50% to 72%

Sites with a large number of processes had an increased DH R/W Station usage

Other sites did not have a significant increase in usage

731-PP-002-001

What were the DH Robotic usage?

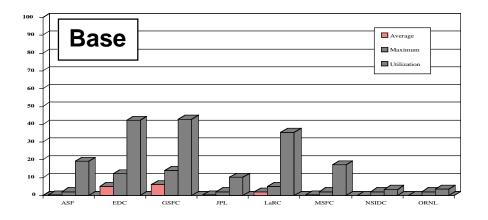


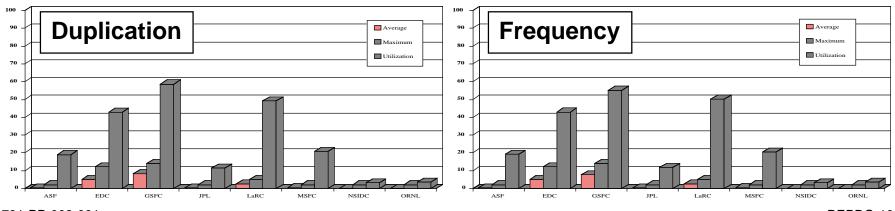
Statistics

- Average use over 21 day run
- Maximum instance

Utilization

- Average / Maximum

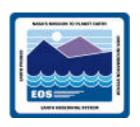




731-PP-002-001

REPRO-10

How did the DH Robots stack up?



Examined those sites with more than two robots

• EDC, GSFC, LaRC

Utilization

• Base 35% to 43%

• Duplication 43% to 59%

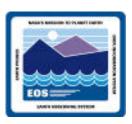
• Frequency 43% to 56%

Sites with a large number of processes had an increased DH Robot usage

Other sites did not have a significant increase in usage

731-PP-002-001

What observations can be drawn?



PGS CPUs

- the demand for processing will increase with reprocessing
- duplication and frequency are similar in that they both double the load on the CPUs and had similar maximum values
- Note: With the associated PGS Disk, frequency however offers a lower peak for maximums and a more distribution of load over the same simulation period as duplication

DH R/W Stations

- average demand for archive access increased in both reprocessing methods when there were many processes being served
- impact did not double, but there was a noticeable increase

DH Robots

- average demand for robot use increased in both reprocessing methods when there were many processes being served
- impact was only 1 or 2 additional robot required

What were the observed values?



PGS CPU	BASE			DUPLICATION			FREQUENCY		
Site ID	Average	Maximum	Utilization	Average	Maximum	Utilization	Average	Maximum	Utilization
ASTER	0.51	2	26%	1.04	2	52%	1.06	2	53%
DAO	1.02	3	34%	2.00	6	33%	2.15	4	54%
DFA/MR	0.00	2	0%	0.00	2	0%	0.00	2	0%
MISR	16.83	40	42%	33.40	40	84%	35.24	40	88%
LIS	0.01	1	1%	0.03	2	1%	0.02	1	2%
MODIS	16.75	36	47%	33.34	36	93%	33.51	36	93%
SWS	0.03	2	2%	0.07	2	3%	0.07	2	3%
MOPITT	0.03	2	2%	0.07	2	3%	0.07	2	4%
CERES TRMM	8.54	20	43%	17.08	20	85%	17.07	20	85%
CERES AM	9.49	22	43%	18.99	22	86%	18.96	22	86%
DH R/W Stations	BASE			DUPLICATION			FREQUENCY		
Site ID	Average	Maximum	Utilization	Average	Maximum	Utilization	Average	Maximum	Utilization
ASF	1.51	3	50%	1.51	3	50%	1.51	3	50%
EDC	19.40	40	48%	20.07	40	50%	20.18	40	50%
GSFC	23.52	48	49%	34.19	48	71%	31.89	48	66%
JPL	0.76	2	38%	0.86	2	43%	0.88	2	44%
LaRC	7.75	16	48%	11.17	16	70%	11.38	16	71%
MSFC	1.33	3	44%	1.57	3	52%	1.55	3	52%
NSIDC	0.24	2	12%	0.24	2	12%	0.24	2	12%
ORNL	0.24	2	12%	0.24	2	12%	0.24	2	12%
DH Robots	BASE			DUPLICATION			FREQUENCY		
Site ID	Average	Maximum	Utilization	Average	Maximum	Utilization	Average	Maximum	Utilization
ASF	0.38	2	19%	0.38	2	19%	0.38	2	19%
EDC	5.05	12	42%	5.14	12	43%	5.16	12	43%
GSFC	5.98	14	43%	8.21	14	59%	7.72	14	55%
JPL	0.21	2	10%	0.23	2	12%	0.24	2	12%
LaRC	1.77	5	35%	2.46	5	49%	2.50	5	50%
MSFC	0.35	2	17%	0.41	2	21%	0.41	2	20%
NSIDC	0.07	2	3%	0.07	2	3%	0.07	2	3%
ORNL	0.07	2	3%	0.07	2	3%	0.07	2	3%